

```
In [3]: import pandas as pd
In [4]: dataset = pd.read_csv('Mail_Customers.csv')
In [5]: dataset.head()
```

```
Out[5]:
```

| | CustomerID | Gender | Age | Annual Income (k\$) | Spending Score (1-100) |
|---|------------|--------|-----|---------------------|------------------------|
| 0 | 1 | Male | 19 | 15 | 20 |
| 1 | 2 | Male | 21 | 15 | 81 |
| 2 | 3 | Female | 70 | 16 | 6 |
| 3 | 4 | Female | 23 | 16 | 77 |
| 4 | 5 | Female | 31 | 17 | 60 |

```
In [6]: dataset.shape
Out[6]: (200, 5)
```

```
In [7]: dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 5 columns):
 #   Column                Non-Null Count  Dtype
---  --
 0   CustomerID            200 non-null    int64
 1   Gender                200 non-null    object
 2   Age                   200 non-null    int64
 3   Annual Income (k$)    200 non-null    int64
 4   Spending Score (1-100) 200 non-null    int64
dtypes: int64(4), object(1)
memory usage: 7.9+ KB
```

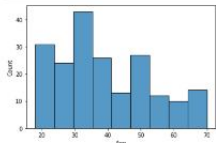
```
In [8]: dataset.describe()
```

```
Out[8]:
```

| | CustomerID | Age | Annual Income (k\$) | Spending Score (1-100) |
|-------|------------|------------|---------------------|------------------------|
| count | 200.000000 | 200.000000 | 200.000000 | 200.000000 |
| mean | 100.500000 | 38.850000 | 60.500000 | 50.200000 |
| std | 57.879385 | 13.968007 | 76.754771 | 75.879377 |
| min | 1.000000 | 18.000000 | 15.000000 | 1.000000 |
| 25% | 50.750000 | 23.750000 | 41.500000 | 34.750000 |
| 50% | 100.500000 | 35.000000 | 61.500000 | 50.000000 |
| 75% | 150.250000 | 49.000000 | 78.000000 | 73.000000 |
| max | 200.000000 | 70.000000 | 117.000000 | 99.000000 |

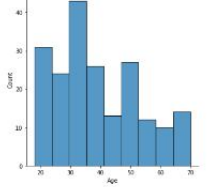
```
'univariate analysis'
```

```
In [9]: from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import confusion_matrix, accuracy_score
In [10]: import seaborn as sns
In [11]: import matplotlib.pyplot as plt
In [12]: sns.histplot(dataset.Age)
plt.show()
```



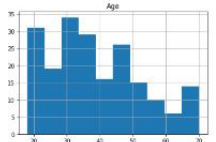
A histogram showing the distribution of Age. The x-axis is labeled 'Age' and ranges from 20 to 70. The y-axis is labeled 'Count' and ranges from 0 to 40. The distribution is roughly bell-shaped, peaking around age 30-40.

```
In [13]: sns.displot(dataset.Age)
Out[13]: <seaborn.axisgrid.FacetGrid at 8x2b6fa0258>
```



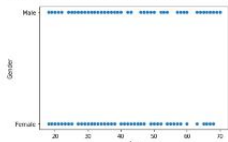
A displot showing the distribution of Age. The x-axis is labeled 'Age' and ranges from 20 to 70. The y-axis is labeled 'Count' and ranges from 0 to 40. The distribution is roughly bell-shaped, peaking around age 30-40.

```
In [14]: dataset.hist('Age')
Out[14]: array([[<AxesSubplot: title='center': 'Age'>]], dtype=object)
```



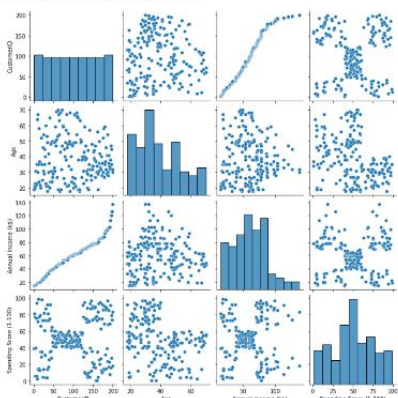
A histogram showing the distribution of Age. The x-axis is labeled 'Age' and ranges from 20 to 70. The y-axis ranges from 0 to 30. The distribution is roughly bell-shaped, peaking around age 30-40.

```
In [15]: sns.scatterplot(x=dataset.Age, y=dataset.Gender)
Out[15]: <AxesSubplot: xlabel='Age', ylabel='Gender'>
```



A scatter plot showing the relationship between Age and Gender. The x-axis is labeled 'Age' and ranges from 20 to 70. The y-axis is labeled 'Gender' and has two categories: 'Male' and 'Female'. The plot shows a dense distribution of points for both genders across the age range.

```
In [16]: sns.pairplot(dataset)
Out[16]: <seaborn.axisgrid.PairGrid at 8xb6f64cfd8>
```



A pair plot showing the relationships between all variables in the dataset: CustomerID, Age, Annual Income (k\$), and Spending Score (1-100). The diagonal shows histograms for each variable. The off-diagonal plots show scatter plots for each pair of variables. The variables are ordered as CustomerID, Age, Annual Income (k\$), and Spending Score (1-100).

```
check the missing values and deals with them
```

```
In [17]: dataset.isna()
```

```
Out[17]:
```

| | CustomerID | Gender | Age | Annual Income (k\$) | Spending Score (1-100) |
|-----|------------|--------|-------|---------------------|------------------------|
| 0 | False | False | False | False | False |
| 1 | False | False | False | False | False |
| 2 | False | False | False | False | False |
| 3 | False | False | False | False | False |
| 4 | False | False | False | False | False |
| ... | ... | ... | ... | ... | ... |
| 195 | False | False | False | False | False |
| 196 | False | False | False | False | False |
| 197 | False | False | False | False | False |
| 198 | False | False | False | False | False |
| 199 | False | False | False | False | False |

200 rows x 5 columns

```
In [ ]:
```